

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
2 May 2002 (02.05.2002)

PCT

(10) International Publication Number  
**WO 02/34771 A2**

(51) International Patent Classification<sup>7</sup>: C07K 14/195

**TETTELIN, Hervé** [BE/US]; The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850 (US).

(21) International Application Number: PCT/GB01/04789

(22) International Filing Date: 29 October 2001 (29.10.2001)

(74) Agents: **HALLYBONE, Huw, George et al.**; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

0026333.5	27 October 2000 (27.10.2000)	GB
0028727.6	24 November 2000 (24.11.2000)	GB
0105640.7	7 March 2001 (07.03.2001)	GB

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(71) Applicants (*for all designated States except US*): **CHIRON S.P.A.** [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT). **THE INSTITUTE FOR GENOMIC RESEARCH** [US/US]; 9712 Medical Center Drive, Rockville, MD 20850 (US).

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **TELFORD, John** [GB/IT]; Chiron S.p.a, Via Fiorentina, 1, I-53100 Siena (IT). **MASIGNANI, Vega** [IT/IT]; Chiron S.p.a, Via Fiorentina, 1, I-53100 Siena (IT). **MARGARIT Y ROS, Immaculada** [IT/IT]; Chiron S.p.a, Via Fiorentina, 1, I-53100 Siena (IT). **GRANDI, Guido** [IT/IT]; Chiron S.p.a, Via Fiorentina, 1, I-53100 Siena (IT). **FRASER, CLAIRE** [US/US]; The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850 (US).

**Published:**

— without international search report and to be republished upon receipt of that report

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: NUCLEIC ACIDS AND PROTEINS FROM STREPTOCOCCUS GROUPS A & B

(57) Abstract: The invention provides proteins from group B streptococcus (*Streptococcus agalactiae*) and group A streptococcus (*Streptococcus pyogenes*), including amino acid sequences and the corresponding nucleotide sequences. Data are given to show that the proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics. The proteins are also targets for

WO 02/34771 A2

ATTGCTCATGAAGTTGATCCAAATAATCAAATCGGATGTATGCTAGCTGCAGGTCAATACTATCCAAATACGTGTATCCACAAGATTATTGGGCT  
TCGATGCAAAAAATAGAGAAAATTTATTTTATTGATGTGCAAGCTCGTGGCAAAATACCTAAATATGCTAAAAAACATTTTGAGCATTAGGT  
ATTTCAATTCAAATGACAGCAGAGATCTTGCTTTGTTAAGAGACTATACGGTTGATTTTATTTTCAATTTCTTACTATTCAAGCCGAGTAGCTTCA  
GGTAATCCAATGCTCAGTGAACAGGTTCAAGAAAATATTTTCGCATCTCTTAAGAATCCTTACTTGAAATCTTCTGAATGGGGATGGCAAAATGAT  
CCGCTTGGATTACGTATTACCTTAAATGCTATCTGGGATCGTTATCAAAAACCGATGTTTATTGTTGAAAATGGAGCTGGAGCAGTAGATATTCT  
GATGAGATGGTTACGTAGAAAGATGATTATCGTATTGATTACCTCCGTCAGCACATTGCTGCAATGAGAGATGCTATATATGTAGATGGTGTAAAT  
CTAATGGCTACTACTACATGGGGATGATTGATTATTAGTATCTGCAGGAACGGTGAAATGGAAAACCGTATGGTTTCATTATGTTGATCGCAAT  
AATAAGGAGAGGGGACACTCAAGCGTTACAAGAAAATCGTTTTATTGTTACAAGAAGGTTATTGCAAGTAATGCGAGTCAAAATAGAA

## SEQ ID 9850

MVKQVFPKGLFWGGATAANQCEGAYNVDRGLANVDVPTGDEDFALISGQKMFDFEEGYFPAKESIDFYHHYKEDLALLAEMGFKTYRMSIAN  
TRIFPKGDELYPNEAGLOFYENIFKBCRKYGIPLVTITHFDCPIYLIKHYGGWRSRKMIGFYERLVRALPTRFKGLVKYWLTFNEINMILHAPFM  
GAGLYFEDGENEQEIKYQAAHHELVASAIAVKLAHEVDNNQIGCMLEAGQYYPNTCHPQDYWASMQKNRENYFFIDVQARGKYPNYAKKHFEHLG  
ISIQMTAEDLALLRDYTVDFISFSYSSRVASGNPTVSEVQENIFASLKNPYLKSEWGWQIDPLGLRITLNAIWDRYQKPMFIVENGLGAVDIP  
DENGIVBDDYRIDYLQHIAAMRDIAIYVDGVNLIGYTTWGCIDLVSAGTGEMEKRYGFIYVDRNNKGEGTLKRYKKSFYWKVVLASNGSQIE

## SEQ ID 9851

ATGGATGGAACCTTTTCTACGCTCAGATAAACTTATGATAAAGCACGGTTTTCTTCTTTATTAACGTTGATGGAGAAGTATGATAITAAATTTGTA  
GCTGCTAGTGGTAATTTATACGATCAATTGTTGTTAACTTTTGGAAATACCCTAATCGTATCGCCTACGCTGAAATGGTGGTGGTGTCTATT  
GACCAAGATGGCACATTACTAAAGAACTTATTTATCTAATGATACAGTTGCAGCTGTTTTATCTATTATATCAGAATCTATCTGAGACACTC  
ATCAGCTTATCTGGTGAAAAACGTTTATACCTTGAGCGAAGAACTCTTATTAATAGGCGGACTGAATTAGAATACTATATGCTTAATTTTATCTAT  
AAGACCATTTGCTGCCCTTATGATGATGACCGCTATTTCCTAATGACACTTTGGGTAAATGAAAATCTTGTTCTGAGATGTTGTTAGATATTAGT  
GAACACTTTAAAAATCATCATATCAGATTGACCTCTAGTGGTTTTGGTTGATTCGATGCTTACCAGCTGATGTCAATAAGCGGATGGTATCGCT  
ATTCTCTTGAAAAGTGGGGATGAAACAGGATCAAGTATGTTTGGTGGTGGGGCAATGATGTGGAATGTTGAGAGCAGCAAAATATTTCT  
TACGCTATGTCAAATGCGCCAGAAGAAATTAAGCTATTGCTAAATATCAACAGTTAGTAACGATCAAGATGGTGTGTTAGAAACGATAGAGAAC  
TTTCTTCAAGATGAGTGATAATGAA

## SEQ ID 9852

MDGTFLRSDKTYDKARFSSLLTLMKYDIKFVAASGNLYDQLLLNFLEYPNRIAYVAENGGRRVIDQDGTLLKETYLNDTVAAVLSYLYQNYPETL  
ISLSGKRSYLERRTPIINRTELEYMFNFYIKDHLPLDDRYFQMTLWVNENLVSEMLLDISEHFKNHHIRLTSSGFGCIDVLPADVKNADGIA  
ILLEKWLKQDQVMVFGDGGNDVEMLRANISYAMSNAPBEEKAIKRYQTVSNDQGVLETIENFLHKMSDNE

## SEQ ID 9853

ATGAAGAAAACCTCTAGTCTATTTTCAATTCCTTATATGCGTTGGTTATTCTTATTGTTATAGCACCTGTGCTTTGATTGCTTGGAACTCGTTC  
TTTGATATTAATGGGCACTTTACTTTAGCTAATTACAGACGTTTTTTAGCTCAGGAACCTATCTTAAATGAGCTTTAACTCGGTCCTATATGCA  
GGCATCGTTAGTTTCAACGCTTTTGATTAGCTATCCAGCTGCTTTTGTCTACAAAATTAACACACAAGCAATATGGCTTAATGCTAGTTATTA  
TTGCCAATGCTGCTTATTTGTTTAAAGCCTATGCTTTTATGGGATATTGTTGTTCAACAGGGGGAATTAACGCTTAACTTATGAGT  
ATCGGCCCCAAAACAGATTCTATTTACAGACTTTTCAATTTCTTTTGTAGCAGCTTACATTGAATTACCGTTTATGTTATTACCAATTTTAAATGCC  
TTGGATGATATTTGATCAAAATCTCATTATGCAAGTATGACTTTGGGTGCAAAATGCATGGCAAACTTTTCAGAAGGTTATATTTCTTTATCTCTG  
AATGGTGTAAAGACAGGTGATCAGTCTGTCTTTATCCCTAGTTTGAATCTCTTATGCTTACGCTCTTATTGGTGGTAAATCGAGTTATCACTTTA  
GGGACAGCCATAGAACCAATTTTAAATCACACAAAATAAAGGTATGGGCTCAACCATAGGGGTTATTTTAAATTTTAGTATGTTGCTATTATG  
TGTTTAAACAAAGGAGAGGAAAGCA

## SEQ ID 9854

MKKTSSLFSIPYMAWLFVLVAPVALIAWNSFFDINGHFTLANYTQTFSSSGTYLKMSFNSVLYAGIVSFIITLLISYPAAYLLTKLKHKLWMLVI  
LPTWINLLKAYAFMGTFGQGGINAFITFIGPKQILFTDFSLFVAAYIELPFMLLPINFALDDIDQNLIVASDDLGANAWQTFQKVIPLSL  
NGVRAGVQSVFIPSLFMLTRLIGGNRVITLGTALIEQHFLITONKMGSGTIGVILILVMVAIMWLTKERKA

## SEQ ID 9855

ATGCTAGAGCAGCTGGGAAACCTCAAACAACCTATCTGCCATTATGTTGTTGGAAACGAATGGTAAAGGATCAACAACCTAGTTATTTACAAAT  
GATTTTACCAATCTGGTTATCAAGTTGGAACATTATACATCACCATTATGTTGACTTTAGAGAGCGTATTAGCATTGATGGTCAAATGATTTCCA  
GAGTCTGATTTTATCAAACTTTGTTGAACTGTTCCGTCAGTTGTTGTAACGGTTGACCTTGAGACCAACCTAGAACAGCTACAGAAATTTGAAGTT  
ATAACAGTTTGTATGTTCTATTTATTTGGCAATAGCTGTCGGGTTGATATCGTCAATTATTGAGGCTGGAATGGGCGGTTATTATGATTCTACCAAC  
ATGTTCAAGGCTTTAGCGGTTACTTGTCGCTCAATGGATTAGATCATCAGGAGGTTTATAGGACGAACTTATGATAGATATAGCTGAGCAAAAAGTA  
GGTGTTTTAAAAAAGGAGTTCCATTCGTATATGCTAATGATCGTCAGGATGTTGAAGAAGTCTTTCAAATAAAGGCCAAGGAAACCCACAGCCAG  
ACTTATCGTTTACACAATGACTTTTATATAAAGAGAGAGGAAATTTTAACTATATAGGTCCTCAAGCAAAATTTGATCATATTAGTTGCAA  
ATGCCGGGACATCAACAAGTTTCCAATGCTAGCATAGCTATACTACCAGTTTATTTAAGAGACAAATATCCAAAGTTAAACCTACAGACCATT  
AAGATGGGCTTGAATGACAAAATGGGTAGGCGGTACAGAAATTAATCTTTCCGAATGTAATGATGATGGCGCTCATAATAATGAAAGTGTGAT  
GCTCTCGTTCAAGTTATAAAAAATATCAGCAGAAAAACGTCATATTCTATTGCTGCAATTAATACTAAACCGATAGAGATGATTAGAAATCA  
TTGCTAGTATCGCTCCGGTTTCGGTTACTAGCTTTGATTACCTTAATCAATAAATCTAGACAAATATCCTAAAGCGGTACACTAGAGTTAGTAT  
TGGAAGAAATGGCTACATGATAAATCTGACATCAGATAAAGATTTCTATGTGATTACTGGCTCCCTTTATTTTATTTCTCAAGTTCTGTCAGGAG  
CTACTATTAATCAAACTACAACAAT

## SEQ ID 9856

MLEQLGNPQWNLISAIHVVGTVNGKGSTISYLQHIPTNSGYQVGTFTSFPYIVDFRERISIDGQMPESDFIKLVETVRPVVERLHLETLNLPATEFEV  
ITVLMFYFNGNSCPVDIVIIRAGMGYYDSTNMFKALAVTSPSISGLDHQEVLRFTYVDLAEQKVGVLKKGVPFVYANDRDQDVEVFQIKAKETHSQ  
TYRLHNDFYIKBENYFNYIGPQANIDHIQLQMPGHEQVSNASIAITSLLLRDKYFKLTLQTIKDGLEMTKWVGRTELIFPNVMIDGAHNESVD  
ALVQVIKQYQKNVHILFAAINTKPIESMLBSLSIAPVSVTSFDYPKSINLDKYFKAYTRVSDHKKWLHDINLTSDDKDFYVITGSLYFISQVRQE  
LLLIKTITT

## SEQ ID 9857

ATGACTATTAATAAGCTTTTATTAGGTTTTGGAACGGTTGCTAAGGGTATTCATATTTGCTAAAAGAAAATCAACATAAGCTACTTTCTTTAGAA  
GGCGAAGATATTGTGATTGATAAAGTATTAGTAAGAGATAATGAAAGCCGCCAGGTTTCATCAATCAGGAGTTTACTTATACTTTGTGACAGAG  
ATAAATACTATCTTCAAGATTCAAAATGATATTGTAGTGGAAATTAATGGGGGGGATTGAGCCAGCTAAAACCTATTTTAGTCAAGCATTAGGA  
TTTGGTAACAATATTGTGACAGCCAAATAAGATCTCATGCTTTACAGAAAAGAGATTGATGGAATTTAGCAGACGCTAGAGGCTAGCTTTATTC  
TATGAGGGAGCAGTTGCTGGAGGCATTCTATTTTAAAGGACCTATCGCATTCGTTTGGCTCAGATAAATGACACGTTATTAGGAATTTCTCAAC  
GGTACCTCCAATCTCATGTTTAACAAAAATGTTTGAAGAGGGATGGTCTTATGAACAAGCTCTAAAAAAGGCACAAGAGTTAGGTTATGCTGAAAGT  
GATCCCAAAATGATGTTGAAGGTATTGATCTGCTACAAAGCCACTATCTTAAGTCAATTTGGATTGGTATGCTTATGATTGTTGATGATGTT  
AATTATAAGGGGATTTCTAGATTTCGCTCAGAGGATGTTGAAGTAGCTCAGGAGATGGGCTTTGCCATTAAGTTGGTAGCTGATCTTCGTGAAACT  
CCAACCTGGTATAAGTTGATAGCTTTCTCCGACCAATTTCTCAAAGCATCCCTTAGCTGAGTAAATCATGATGATGAATGATGATGATGATGAA  
TCAATAGGGAATGGTCACTCTTTTATGAGACAGGTGCGGGACAAAATCCAACAGCAACCTCTGTTTATAGCGGATATCATGATATTAGTCGT

RN - 440183-53-9      REGISTRY  
FS - PROTEIN SEQUENCE  
SQL - 478

Sequence	Patent
Source	Reference

Not Given|WO2002034771  
|claimed SEQID  
|9850

```

SEQ      1  MVKQVFPKGF  LWGGATAANQ  CEGAYNVDGR  GLANVDVVP  GEDRFALISG
      51  QKRMDFEEG  YFYPAKESID  FYHHYKEDLA  LLAEMGFKTY  RMSIAWTRIF
     101  PKGDELYPNE  AGLOFYENIF  KECRKYGIEP  LVTITHFDCP  IYLIKHYGGW
     151  RSRKMIGFYE  RLVRALFTRF  KGLVKYWLTF  NEINMILHAP  FMGAGLYFED
     201  GENQEIQKYQ  AAHELVASA  IAVKIAHEVD  PNNQIGCMLA  AGQYYPNTCH

      251  " " " "  PODYWASMOX  NRENYFFIDV  QARGKYPNYA  KKHFEHLGIS  IQMTAEDLAL

     301  LRDTYVDFIS  FSYYSRVAS  GNPTVSEQVQ  ENIFASLKNP  YLKSSEWGWQ
     351  IDPLGLRITL  NAIWDRYQKP  MFIVENGLGA  VDIPDENGIV  EDDYRIDYLR
     401  QHIAAMRDAI  YVDGVNLIGY  TTWGCIDLVS  AGTGEMEKRY  GFYIVDRNNK
     451  GEGTLKRYKK  KSFYWKYKVI  ASNGSOIE

```

**HITS AT: 244-276**